



PATENT ABSTRACTS OF JAPAN

(11) Publication number: **08221093 A**

(43) Date of publication of application: 30 . 08 . 96

(51) Int. Cl. **G10L 3/02**
G10L 3/00

(21) Application number: 07029336

(71) Applicant: **SONY CORP**

(22) Date of filing: 17 . 02 . 95

(72) Inventor: **CHIYAN JIYOSUFU**

(54) METHOD OF NOISE REDUCTION IN VOICE SIGNAL

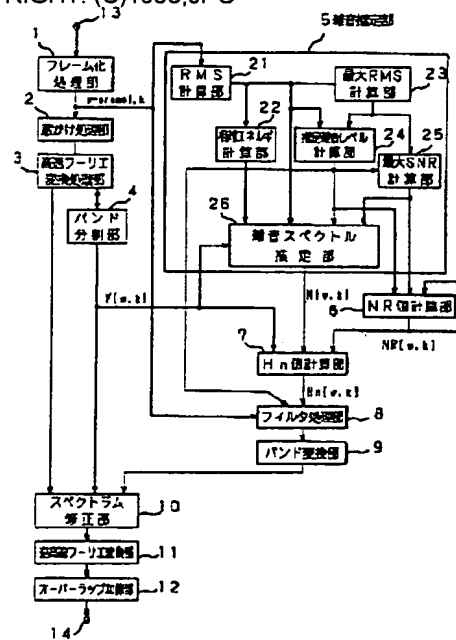
(57) Abstract:

PURPOSE: To provide a method of reducing noise in voice signal which is capable of simplifying computation in suppressing noise in input signals.

CONSTITUTION: The method is conducted by a fast Fourier transform processing section 3 which transforms input voice signals to the signals of frequency axis, an Hn-value calculating section 7 which controls the filter characteristics of filter processing in the case of eliminating noise part from input voice signals, and a spectrum correcting section 10 which reduces noise from input voice signals by the filter processing responding to the filter characteristics obtained in the Hn-value calculating section 7. The Hn-value calculating section 7 calculates the Hn-value in accordance with the maximum SN ratio of the input signal spectrum per frame obtained by the fast Fourier transform processing section 3 and the value obtained from the estimated noise level, and according to this Hn-value, the processing to eliminate noise in the spectrum correcting

section 10 is controlled.

COPYRIGHT: (C)1996,JPO



(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平 8-221093

(43) 公開日 平成 8 年 (1996) 8 月 30 日

(51) Int. Cl.	識別記号	庁内整理番号	F I	技術表示箇所
G10L 3/02	301			
3/00	513			

G10L 3/02 301 D
3/00 513 A

審査請求 未請求 請求項の数 4 (全 14 頁)

(21) 出願番号 特願平 7-29336

(22) 出願日 平成 7 年 (1995) 2 月 17 日

(71) 出願人 000002185

ソニー株式会社

東京都品川区北品川 6 丁目 7 番 35 号

(72) 発明者 チャン・ジョセフ

東京都品川区北品川 6 丁目 7 番 35 号 ソニー株式会社内

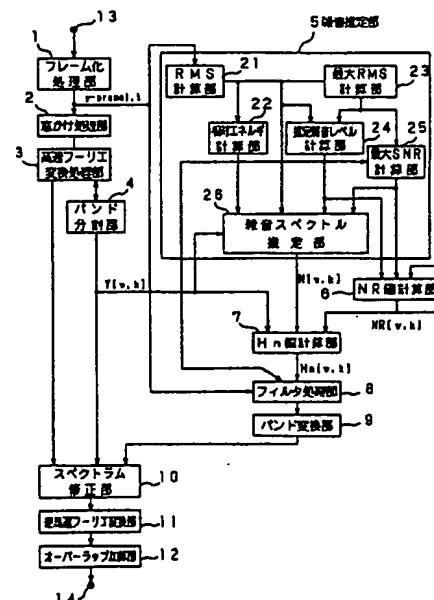
(74) 代理人 弁理士 小池 晃 (外 2 名)

(54) 【発明の名称】 音声信号の雑音低減方法

(57) 【要約】

【目的】 入力信号の雑音抑圧を行うのに演算を簡略化することが可能である音声信号の雑音低減方法を提供する。

【構成】 入力音声信号を周波数軸の信号に変換する高速フーリエ変換処理部 3 と、上記入力音声信号から雑音部分を除去する際のフィルタ処理のフィルタ特性を制御する H_n 値計算部 7 と、上記 H_n 値計算部 7 にて得られるフィルタ特性に応じたフィルタ処理により、上記入力音声信号から雑音を低減するスペクトラム修正部 10 とを有して成り、 H_n 値計算部 7 は、高速フーリエ変換処理部 3 にて得られる入力信号スペクトルのフレーム毎の最大 S/N 比及び推定雑音レベルとから得られる値に応じて H_n 値を算出し、この H_n 値に応じて、上記スペクトラム修正部 10 における雑音を除去する処理を制御する。



【特許請求の範囲】

【請求項 1】 入力音声信号から雑音除去することで雑音抑圧を行う音声信号の雑音低減方法であって、上記入力音声信号を周波数軸の信号に変換する変換工程と、

上記入力音声信号から雑音部分を除去する際のフィルタ処理のフィルタ特性を制御する制御工程と、

上記制御工程にて得られるフィルタ特性に応じたフィルタ処理により、上記入力音声信号から雑音を低減する雑音低減工程とを有し、

上記制御工程は、上記変換工程にて得られる入力信号スペクトルのレベルと上記入力信号スペクトルに含まれる推定雑音スペクトルのレベルとの比に基づいて求められる第一の値と、上記入力信号スペクトルのフレーム毎の信号レベルと推定雑音レベルとの比の最大値及び推定雑音レベルから求められる第二の値とにより上記フィルタ特性を制御する工程であることを特徴とする音声信号の雑音低減方法。

【請求項 2】 上記第一の値は、予め設定された入力信号スペクトルのレベルと推定雑音スペクトルのレベルとで構成される表から得られる値を用いて求められることを特徴とする請求項 1 記載の音声信号の雑音低減方法。

【請求項 3】 上記第二の値は、上記信号レベルと推定雑音レベルとの比の最大値及びフレーム毎の雑音レベルに応じて得られる値であると共に、上記フィルタ特性に応じたフィルタ処理による最大雑音低減量を dB 領域で略線形的に変化させるように調節する値であることを特徴とする請求項 1 記載の音声信号の雑音低減方法。

【請求項 4】 上記推定雑音レベルは、上記フレーム毎の入力信号の振幅の 2 乗平均値の平方根と 2 乗平均値の平方根の最大値とに基づいて得られる値であり、上記信号レベルと推定雑音レベルとの比の最大値は、上記推定雑音レベル及び上記 2 乗平均値の平方根の最大値に基づいて算出される値であり、

上記 2 乗平均値の平方根の最大値は、上記フレーム毎の入力信号の振幅の 2 乗平均値の平方根と、一つ前のフレームの 2 乗平均値の平方根の最大値とに基づいて得られる値と所定の値との最大値であることを特徴とする請求項 1 記載の音声信号の雑音低減方法。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、入力音声信号から雑音除去することで雑音抑圧を行う音声信号の雑音低減方法に関する。

【0002】

【従来の技術】 携帯用電話機や音声認識等の応用において、収音された音声信号に含まれる環境雑音や背景雑音等の雑音を抑圧し、音声 (speech) 成分を強調することが必要とされている。

【0003】 このような音声強調、あるいは雑音低減の

技術として、減衰ファクタの調整のために条件付き確率関数を用いる例が、文献「軟判定雑音抑圧フィルタを用いる音声強調」(Speech Enhancement Using a Soft-Decision Noise Suppression Filter, R.J.McAulay, M.L.Malpass, IEEE Trans. Acoust., Speech, Signal Processing, Vol.28, pp.137-145, April 1980)や、「移動電話システムにおける周波数領域雑音抑圧研究」(Frequency Domain Noise Suppression Approach in Mobile Telephone Systems, J.Yang, IEEE ICASSP, Vol.11, pp.363-366, April 1993)等に開示されている。

【0004】

【発明が解決しようとする課題】 しかしながら、これらの雑音抑圧技術においては、不適切な固定の SNR (信号対雑音比) に基づく動作をするため、または不適切な抑圧フィルタにより、音色の不自然さや歪んだ音声を生ずることがある。実際の運用の際に、最適の性能を得ために雑音抑圧装置のパラメータの 1 つである SNR を調整することは、ユーザにとって望ましいことではない。さらに、従来の音声信号強調技術は、短時間 SNR の大きな変動がある音声信号に対して副作用として発生してしまう歪を持たせずに充分雑音を除去することは困難である。

【0005】 また、このような音声強調、あるいは雑音低減方法においては、雑音区間検出の技術が用いられ、入力レベルやパワー等を所定の閾値と比較することにより、雑音区間判別を行っているが、音声 (speech) にトラッキングすることを防ぐために閾値の時定数を大きくすると、ノイズレベルが変化するとき、特に増加するときに従従できなくなり、誤判別が生じ易くなる。

【0006】 ここで、本発明人は、上述した問題を解決するために、特願平 6 - 9 9 8 6 9 号において、音声信号の雑音低減方法を提案している。

【0007】 上記音声信号の雑音低減方法は、入力音声信号に基づいて算出された SN 比及び音声存在確率に基づいて、音声 (speech) 成分を算出するための最尤フィルタを適応的に制御することで雑音抑圧を行う音声信号の雑音低減方法であって、上記音声存在確率の算出に、入力信号のスペクトルから推定雑音スペクトルを減算したものをを用いることを特徴とするものである。

【0008】 また、上記音声信号の雑音低減方法によれば、上記最尤フィルタが上記入力音声信号の SN 比に応じて最適の抑圧フィルタに調整されるため、上記入力音声信号に対して充分な雑音除去を行うことが可能である。

【0009】 ところが、上記音声存在確率を算出するのに、複雑な演算を有すると共に、膨大な演算量が要求されるため、演算の簡略化が望まれる。

【0010】 そこで、本発明は、上述した実情に鑑みてなされたものであり、入力信号の雑音抑圧を行うのに演算を簡略化することが可能である音声信号の雑音低減方

法を提供することを目的とする。

【0011】

【課題を解決するための手段】本発明に係る音声信号の雑音低減方法は、上述した問題を解決するために、入力音声信号から雑音除去することで雑音抑圧を行う音声信号の雑音低減方法であって、上記入力音声信号を周波数軸の信号に変換する変換工程と、上記入力音声信号から雑音部分を除去する際のフィルタ処理のフィルタ特性を制御する制御工程と、上記制御工程にて得られるフィルタ特性に応じたフィルタ処理により、上記入力音声信号から雑音を低減する雑音低減工程とを有し、上記制御工程を、上記変換工程にて得られる入力信号スペクトルのレベルと上記入力信号スペクトルに含まれる推定雑音スペクトルのレベルとの比に基づいて求められる第一の値と、上記入力信号スペクトルのフレーム毎の信号レベルと推定雑音レベルとの比の最大値及び推定雑音レベルとから求められる第二の値とにより上記フィルタ特性を制御する工程とするものである。

【0012】また、本発明は、上記音声信号の雑音低減方法において、上記第一の値は、予め設定された入力信号スペクトルのレベルと推定雑音スペクトルのレベルとで構成される表から得られる値を用いて求められるものである。

【0013】また、本発明は、上述の音声信号の雑音低減方法において、上記第二の値は、上記信号レベルと推定雑音レベルとの比の最大値及びフレーム毎の雑音レベルに応じて得られる値であると共に、上記フィルタ特性に応じたフィルタ処理による最大雑音低減量をdB領域で略線形的に変化させるように調節する値であるものである。

【0014】また、本発明は、上述の音声信号の雑音低減方法において、上記推定雑音レベルは、上記フレーム毎の入力信号の振幅の2乗平均値の平方根と2乗平均値の平方根の最大値とに基づいて得られる値であり、上記信号レベルと推定雑音レベルとの比の最大値は、上記推定雑音レベル及び上記2乗平均値の平方根の最大値に基づいて算出される値であり、上記2乗平均値の平方根の最大値は、上記フレーム毎の入力信号の振幅の2乗平均値の平方根と、一つ前のフレームの2乗平均値の平方根の最大値とに基づいて得られる値と所定の値との最大値であるものである。

【0015】

【作用】本発明に係る音声信号の雑音低減方法によれば、第一の値は、変換工程にて入力音声信号から得られる入力信号スペクトルと、この入力信号スペクトル中に含まれる雑音推定スペクトルとの比に基づいて算出される値であると共に、雑音低減処理のフィルタ処理における雑音低減量を決定するフィルタ特性の初期値を設定する。また、第二の値は、上記入力信号スペクトルの信号レベルと推定雑音レベルとの比の最大値いわゆる最大S

N比及び推定雑音レベルに基づいて算出される値であると共に、上記最大SN比に応じて上記フィルタ特性を可変制御する値である。また、上記第一の値と上記第二の値とで可変制御されるフィルタ特性に応じたフィルタ処理により、上記入力音声信号から上記最大SN比に応じた雑音低減量で雑音が除去される。

【0016】また、本発明によれば、第一の値を求めるのに、予め設定された入力信号スペクトルのレベルと推定雑音スペクトルのレベルとで構成される表を用いることが可能であるため、演算量の削減化を図ることができる。

【0017】また、本発明によれば、上記第二の値は、上記最大SN比及びフレーム毎の雑音レベルに応じて得られる値であるため、上記フィルタ処理による最大雑音低減量を上記最大SN比に応じて、dB領域で略線形的に変化させるように上記フィルタ特性を調節することができる。

【0018】

【実施例】以下、本発明に係る音声信号の雑音低減方法について、図面を参照しながら詳細に説明する。

【0019】ここで、ノイズ低減装置に適用した本発明の音声信号の雑音低減方法の一例を図1に示す。

【0020】上記ノイズ低減装置は、入力音声信号を周波数軸の信号に変換する高速フーリエ変換処理部3と、上記入力音声信号から雑音部分を除去する際のフィルタ処理のフィルタ特性を制御するHn値計算部7と、上記Hn値計算部7にて得られるフィルタ特性に応じたフィルタ処理により、上記入力音声信号から雑音を低減するスペクトラム修正部10とを有して成るものである。

【0021】上記ノイズ低減装置において、音声信号入力端子13から入力される入力音声信号 $y[t]$ は、フレーム化処理部1に送られる。このフレーム化処理部1からの出力であるフレーム化信号 $y-frame_{j,k}$ は、窓かけ処理部2、雑音推定部5内の2乗平均値の平方根(RMS: root mean square)計算部21及びフィルタ処理部8に送られる。

【0022】窓かけ処理部2からの出力は、高速フーリエ変換処理部3に送られる。また、高速フーリエ変換処理部3からの出力は、スペクトル修正部10に送られる。他に、バンド分割部4にも送られる。バンド分割部4からの出力は、上記スペクトル修正部10、上記雑音推定部5内の雑音スペクトル推定部26及びHn値計算部7に送られる。また、スペクトル修正部10からの出力は、逆高速フーリエ変換処理部11及びオーバーラップ加算部12を介して、音声信号出力端子14に送られる。

【0023】また、RMS計算部21からの出力は、相対エネルギー計算部22、最大RMS計算部23、推定雑音レベル計算部24及び雑音スペクトル推定部26に送られる。さらに、上記最大RMS計算部23からの出力は、推定雑音レベル計算部24及び最大SN比計算部2

10

20

30

40

50

5 に送られる。また、相対エネルギー計算部 22 からの出力は、上記雑音スペクトル推定部 26 に送られる。推定雑音レベル計算部 24 からの出力は、フィルタ処理部 8、最大 SN 比計算部 25、雑音スペクトル推定部 26 及び NR 値計算部 6 に送られる。また、最大 SN 比計算部 25 からの出力は、NR 値計算部 6 及び上記雑音スペクトル推定部 26 に送られる。また、雑音スペクトル推定部 26 からの出力は、Hn 値計算部 7 に送られる。

【0024】また、NR 値計算部 6 からの出力は、再度 NR 値計算部 6 に送られる他、上記 Hn 値計算部 7 にも送られる。

【0025】Hn 値計算部 7 からの出力は、フィルタ処理部 8 及びバンド変換部 9 を介して、上記スペクトル修正部 10 に送られる。

【0026】以下、上記ノイズ低減装置の第一の例の動作を説明する。

【0027】音声信号入力端子 13 には、音声 (Speech) 成分と雑音 (Noise) 成分とを含む入力音声信号 $y[t]$ が供給されている。この入力音声信号 $y[t]$ は、例えば標準化周波数が FS のデジタル信号であり、フレーム化処理部 1 に送られて、フレーム長が FL

$$W_{input}[j] = \left(\frac{1}{2} - \frac{1}{2} \cdot \cos \left(\frac{2 \cdot \pi \cdot j}{FL} \right) \right)^{\frac{1}{4}} \quad 0 \leq j \leq FL \quad \dots (1)$$

$$W_{output}[j] = \left(\frac{1}{2} - \frac{1}{2} \cdot \cos \left(\frac{2 \cdot \pi \cdot j}{FL} \right) \right)^{\frac{3}{4}} \quad 0 \leq j \leq FL \quad \dots (2)$$

【0030】次に、高速フーリエ変換処理部 3 では、256 ポイントの高速フーリエ変換処理が施され、得られた周波数スペクトル振幅値は、バンド分割部 4 により、例えば 18 バンドに分割される。これらの各バンドの周波数レンジの一例を、次の表 1 に示す。このバンド分割された周波数スペクトルの振幅値は、入力信号スペクトルの振幅 $Y[w, k]$ となり、上述したように、各部に出力される。

【0031】

【表 1】

バンド番号	周波数範囲
0	0~125 Hz
1	125~250 Hz
2	250~375 Hz
3	375~563 Hz
4	563~750 Hz
5	750~938 Hz
6	938~1125 Hz
7	1125~1313 Hz
8	1313~1563 Hz
9	1563~1813 Hz
10	1813~2063 Hz
11	2063~2313 Hz
12	2313~2563 Hz
13	2563~2813 Hz
14	2813~3063 Hz
15	3063~3375 Hz
16	3375~3688 Hz
17	3688~4000 Hz

サンプルのフレームに分割され、以下各フレーム毎に処理が行われる。このフレームの時間軸方向の移動量であるフレーム間隔は FI サンプルであり、第 (k+1) フレームは第 k フレームから FI サンプル後に開始されることになる。また、上記周波数やサンプル数の具体例を挙げると、標準化周波数 FS を 8000、または 8 kHz とするとき、フレーム間隔 FI を 80 サンプルとすると 10 ms に、また、フレーム長 FL を 160 サンプルとすると 20 ms に相当することになる。

【0028】窓かけ処理部 2 では、次の直交変換である、例えば高速フーリエ変換処理部 2 での計算に先立って、上記フレーム化処理部 1 より送られる各フレーム化信号 $y-frame_{j,k}$ に対して、窓関数 w_{input} による窓かけ処理が施される。なお、各フレーム毎の信号処理の終段での後述する逆高速フーリエ変換処理のあとには、出力信号に対して窓関数 w_{output} による窓かけ処理が施される。このような各窓関数 w_{input} 及び w_{output} の一例を、次の (1) 式及び (2) 式にそれぞれ示す。

【0029】

【数 1】

【0032】これらの周波数帯域は、人間の聴覚システ

ムが高域ほど知覚分解能が劣化することに基づいている。各帯域の振幅として、対応する周波数レンジ内の最大FFT（高速フーリエ変換処理における周波数帯域の）振幅を用いる。

【0033】次に、雑音推定部5においては、フレーム化信号 $y-framej,k$ のノイズが音声（speech）から区別され、ノイズと推定されるフレームが検出されると共に、推定雑音レベル値と、最大SN比とがNR値計算部6に送られている。この雑音区間推定、あるいはノイズフレーム検出処理は、例えば3種類の検出処理を組み合わせている。この雑音区間推定の具体例について説明する。

【0034】RMS計算部21では、上記各フレーム誤毎の信号のRMS値の計算が行われ、出力される。第kフレームのRMS値である $RMS[k]$ は、次式で計算される。

$$dB_{rel}[k] = 10 \log_{10} \left(\frac{E_{decay}[k]}{E[k]} \right) \quad \dots(4)$$

$$E[k] = \sum_{n=1}^{FL} (y-framej,k)^2 \quad \dots(5)$$

$$E_{decay}[k] = \max \left(E[k], \left(\exp \left(\frac{-FL}{0.65 \cdot FS} \right) \right) \cdot E_{decay}[k-1] \right) \quad \dots(6)$$

【0038】ここで、上記（5）式は、 $FL \cdot (RMS[k])^2$ と表すことができるが、RMS計算部21での上記（3）式の計算の途中で得られる上記（5）式の値をそのまま相対エネルギー計算部21に送るようにしてもよいことは勿論である。また、上記（6）式においては、減衰時間（ディケイタイム）を0.65秒とした例を示している。

【0039】このような、エネルギー $E[k]$ 及び減衰エネルギー $E_{decay}[k]$ の具体例を、図2に示す。

【0040】最大RMS計算部23では、後述する推定雑音レベル値と信号レベルと推定雑音レベルとの比の最大値いわゆる最大SN比とを見積もるのに必要な最大RMS値が求められ、出力される。この最大RMS値 $MaxRMS[k]$ は、以下の（7）式にて算出される。

（7）式で、 θ は減衰定数（decay constant）であり、例えば3.2秒で最大RMS値が $1/e$ だけ減衰するような値、すなわち $\theta = 0.993769$ が用いられる。

【0041】

【数4】

$$MaxRMS[k] = \max(4000, RMS[k], \theta \cdot MaxRMS[k-1] + (1-\theta) \cdot RMS[k]) \quad \dots(7)$$

【0042】推定雑音レベル計算部24では、バックグラウンドノイズ、あるいは背景雑音のレベルを評価するのに最適な最小のRMS値が求められ、出力される。この推定雑音レベル値 $MinRMS[k]$ は、現時点から前に5個の局所極小値（local minimum）、すなわち

【0035】

【数2】

$$RMS[k] = \sqrt{\frac{1}{FL} \cdot \sum_{j=0}^{FL-1} (y-framej,k)^2} \quad \dots(3)$$

【0036】相対エネルギー計算部22では、前フレームからの減衰エネルギーに関連する第kフレームの相対エネルギーを示す $dB_{rel}[k]$ が計算され、得られた値が出力されている。このdB表示の相対エネルギーdB

$rel[k]$ は、次の（4）式により計算され、この（4）式中のエネルギー値 $E[k]$ 及び減衰エネルギー値 $E_{decay}[k]$ は、それぞれ次の（5）式及び（6）式により求められる。

【0037】

【数3】

（8）式を満たす値の中で最小となる値である。

【0043】

【数5】

$$\begin{aligned} (RMS[k] < 0.6 \cdot MaxRMS[k]) & \text{ かつ} \\ RMS[k] < 4000 & \text{ かつ} \\ RMS[k] < RMS[k+1] & \text{ かつ} \\ RMS[k] < RMS[k-1] & \text{ かつ} \\ RMS[k] < RMS[k-2] & \text{ 又は} \\ (RMS[k] < MinRMS) & \dots(8) \end{aligned}$$

【0044】この推定雑音レベル値 $MinRMS[k]$ は、音声（Speech）無しの背景雑音、いわゆるバックグラウンドノイズのときに上昇してゆくように設定されている。ノイズレベルが高いときのノイズレートの指数関数的であるが、低いノイズレベルのときには、より大きな上昇を得るために固定の上昇レートが用いられる。

【0045】これらのRMS値 $RMS[k]$ 、推定雑音レベル値 $MinRMS[k]$ 及び最大RMS値 $MaxRMS[k]$ の具体例を、図3に示す。

【0046】また、最大SN比計算部25では、上記最大RMS値及び上記推定雑音レベル値を用いて、以下の（9）式により最大SN比が推定され、この最大SN比 $MaxSNR[k]$ が算出され、出力されている。

【0047】

【数6】

$$MaxSNR[k] = 20 \cdot \log_{10} \left(\frac{MaxRMS[k]}{MinRMS[k]} \right) - 1 \quad \dots(9)$$

【0048】また、この最大SN比値MaxSNRからは、相対ノイズレベルを示す0から1までの範囲の正規

$$NR_level[k] = \begin{cases} \left(\frac{1}{2} + \frac{1}{2} \cos \left(\pi \cdot \frac{MaxSNR[k] - 30}{20} \right) \right) \times \\ \quad \left(1 - 0.002(MaxSNR[k] - 30)^2 \right) & 30 < MaxSNR[k] \leq 50 \quad \dots(10) \\ 0.0 & MaxSNR[k] > 50 \\ 1.0 & \text{上記以外} \end{cases}$$

【0050】次に、雑音スペクトル推定部26での動作を説明する。上記相対エネルギー計算部22、推定雑音レベル計算部24及び最大SN比計算部25において算出した値は、音声(speech)を背景雑音(background noise)から区別するために用いられる。次の条件が正しいとき、第kフレーム中の信号は背景雑音として分類され

$$\begin{aligned} & ((RMS[k] < NoiseRMS_{thres}[k]) \text{ 又は } \\ & (dB_{rel}[k] > dB_{thres}[k])) \quad \text{かつ} \\ & (RMS[k] < RMS[k-1] + 200) \quad \dots(11) \end{aligned}$$

ここで $NoiseRMS_{thres}[k] = 1.05 + 0.45 \cdot NR_level[k] \times MinRMS[k]$

$$dB_{thres,rel}[k] = \max(MaxSNR[k] - 4.0, 0.9 \cdot MaxSNR[k])$$

【0052】ここで図4は、上記(11)式中のdB表示の相対エネルギーdB_{rel}[k]と、最大SN比MaxSNR[k]と、雑音判別の閾値の1つであるdB_{thres,rel}[k]との具体例を示している。

【0053】また図5は、上記(10)式中のMaxSNR[k]の関数としてのNR_level[k]を示している。

【0054】第kフレームが背景雑音、あるいはノイズとして分類される場合、上記ノイズスペクトルの時間平均推定値N[w, k]は、現在フレームの信号の入力信号スペクトルの振幅Y[w, k]によって、次の(12)式のように更新される。なお、wは上記バンド分割のバンド番号を示すものである。

【0055】

【数9】

$$NR[w, k] = \begin{cases} adj[w, k] NR[w, k-1] - \delta_{NR} < adj[w, k] \\ < NR[w, k-1] + \delta_{NR} \\ NR[w, k-1] - \delta_{NR} NR[w, k-1] - \delta_{NR} \geq adj[w, k] \\ NR[w, k-1] + \delta_{NR} NR[w, k-1] + \delta_{NR} \leq adj[w, k] \end{cases} \quad \dots(13)$$

$$\delta_{NR} = 0.004$$

$$adj[w, k] = \min(adj1[k], adj2[k]) - adj3[w, k] \quad \dots(14)$$

化パラメータNR_levelが算出される。このNR_levelには、以下の関数が用いられる。

【0049】

【数7】

る。こうして分類された背景雑音が示す振幅値は、ノイズスペクトルの時間平均推定値N[w, k]として算出され、出力される。

【0051】

【数8】

$$N[w, k] = \alpha \cdot \max(N[w, k-1], Y[w, k]) + (1 - \alpha) \cdot \min(N[w, k-1], Y[w, k]) \quad \dots(12)$$

$$\alpha = \exp \left(\frac{-FI}{0.5 \cdot FS} \right)$$

【0056】ここで、第kフレームが音声(speech)として分類された場合、N[w, k]はN[w, k-1]の値をそのまま用いる。

【0057】NR値計算部6では、急激にフィルタ応答が変化することを回避するために用いる値であるNR[w, k]を計算し、得られたNR[w, k]値が出力される。このNR[w, k]は、0から1の大きさの値であり、(13)式にて定義される値である。

【0058】

【数10】

【0059】また、(13)式中のadj[w, k] 50 は、後述する効果を考慮したパラメータであって、(1

4) 式にて定義される。

【0060】ここで、(14) 式中の $adj1[k]$ は、全ての帯域において、高い SN 比における後述するフィルタ処理による雑音抑圧動作を抑える効果を有する

$$adj1[k] = \begin{cases} 1 & MaxSNR[k] < 29 \\ 1 - \frac{MaxSNR[k] - 29}{14} & 29 \leq MaxSNR[k] < 43 \quad \dots(15) \\ 0 & \text{上記以外} \end{cases}$$

【0062】また、(14) 式中の $adj2[k]$ は、非常に低いノイズレベルや非常に高いノイズレベルに対して、上記フィルタ処理による雑音抑圧レートを抑える効果を有する値であり、以下の(16)式にて定義され

$$adj2[k] = \begin{cases} 0 & MinRMS[k] < 20 \\ \frac{MinRMS[k] - 20}{40} & 20 \leq MinRMS[k] < 60 \\ 1 & 60 \leq MinRMS[k] < 1000 \quad \dots(16) \\ 1 - \frac{(MinRMS[k] - 1000)}{1000} & 1000 \leq MinRMS[k] < 1800 \\ 0.2 & MinRMS[k] \geq 1800 \end{cases}$$

【0064】また、(14) 式中の $adj3[w, k]$ は、2375 Hz から 4000 Hz の間での 18 dB から 15 dB への最大雑音低減量を抑える効果を有する値であり、以下の(17)式にて定義される。

【0065】
【数13】

$$adj3[w, k] = \begin{cases} 0 & w < 2375 \text{ Hz} \\ \frac{0.059415(w - 2375)}{4000 - 2375} & \text{上記以外} \quad \dots(17) \end{cases}$$

【0066】なお、上述した値である $NR[w, k]$ と、最大雑音低減量 (dB) との関係は、図 6 に示すように、dB 領域において略線形であることがわかる。

$$Hn[w, k] = 1 - (2 \cdot NR[w, k] - NR^2[w, k]) \cdot (1 - H[w][S/N = \gamma]) \quad \dots(18)$$

$$H[w][S/N = \gamma] = \frac{1}{2} \left(1 + \sqrt{1 - \frac{1}{x^2[w, k]}} \right) \cdot P(H1|Y_w)[S/N = \gamma] + G_{min} \cdot P(H0|Y_w)[S/N = \gamma] \quad \dots(19)$$

【0069】また、上記(18) 式中の値 $H[w][S/N = r]$ は、SN 比をある値 r に固定したとき最適なノイズ抑圧フィルタ特性に当たり、(19) 式にて求められる値である。また、この値は、予め求めることができ、 $Y[w, k]/N[w, k]$ の値に応じてテーブル化することが可能な値である。なお、(19) 式中の $x[w, k]$ は $Y[w, k]/N[w, k]$ に相当し、 G_{min} は $H[w][S/N = r]$ の最小利得を示すパラメータである。また、 $P(H1|Y_w)[S/N = r]$

値であり、以下の(15) 式にて定義される。

【0061】
【数11】

【0067】 Hn 値計算部 7 は、バンド分割された入力信号スペクトルの振幅 $Y[w, k]$ と、ノイズスペクトルの時間平均推定値 $N[w, k]$ と、上記 $NR[w, k]$ とから、上記バンド分割された入力信号スペクトルの振幅 $Y[w, k]$ から雑音成分を低減するためのプレフィルタである。ここでは、 $Y[w, k]$ が $N[w, k]$ に応じて $Hn[w, k]$ に変換され、このフィルタ応答 $Hn[w, k]$ が出力される。なお、この $Hn[w, k]$ 値は、以下の(18) 式に基づいて算出される。

【0068】
【数14】

及び $P(H0|Y_w)[S/N = r]$ は、各入力信号スペクトルの振幅 $Y[w, k]$ の状態を示すパラメータであり、 $P(H1|Y_w)[S/N = r]$ は $Y[w, k]$ に音声 (speech) 成分と雑音成分とが混在した状態を指し、 $P(H0|Y_w)[S/N = r]$ は $Y[w, k]$ に雑音成分のみが含まれる状態を指すパラメータである。また、これら値は、以下の(20) 式にて算出される。

【0070】
【数15】

$$P(H1|Y_w)_{[S/N=r]} = 1 - P(H0|Y_w)_{[S/N=r]} =$$

$$\frac{P(H1) \cdot (\exp(-\gamma^2)) \cdot I_0(2 \cdot \gamma \cdot x[w, k])}{P(H1) \cdot (\exp(-\gamma^2)) \cdot I_0(2 \cdot \gamma \cdot x[w, k]) + P(H0) \cdot (\exp(-x^2[w, k]))}$$

$$\dots (20)$$

$$P(H1) = P(H0) = 0.5$$

【0071】(20)式によれば、 $P(H1|Y_w)$ $[S/N=r]$ 及び $P(H0|Y_w)$ $[S/N=r]$ は $x[w, k]$ の関数であることがわかる。また、 $I_0(2 \cdot r \cdot x[w, k])$ は、ベッセル関数であり、 r と $x[w, k]$ との値に応じて求められる。なお、 $P(H1)$ 及び $P(H0)$ は、共に 0.5 に固定される。このように、パラメータを単純化することで、演算量を従来の略 5 分の 1 に削減することができる。

【0072】また、この H_n 値計算部 7 にて得られる $H_n[w, k]$ 値と $x[w, k]$ 、すなわち $Y[w, k]/N[w, k]$ 比との関係は、図 7 に示すように、 $Y[w, k]/N[w, k]$ が高いところ、すなわち雑音成分よりも音声 (speech) 成分の方が多くなるところにおいて、 $H_n[w, k]$ 値が大きくなり、すなわち抑圧が弱くなり、また、 $Y[w, k]/N[w, k]$ が低いところ、すなわち雑音成分よりも音声 (speech) 成分の方が少なくなるところにおいて、 $H_n[w, k]$ 値が小さくなる、すなわち抑圧が強くなる関係である。なお、上記関係において、 $r=2.7$ 、 $G_{min}=-18\text{dB}$ 、 $NR[w, k]=1$ のときの曲線を実線で示している。

$$\text{Step1: } H1[w, k] = \max(\text{median}(Hn[w-1, k], Hn[w, k], Hn[w+1, k]), Hn[w, k]) \dots (21)$$

但し、 $(w-1)$ 、又は $(w+1)$ が存在しないとき 30 は、 $H1[w, k] = Hn[w, k]$

$$\text{Step2: } H2[w, k] = \min(\text{median}(H1[w-1, k], H1[w, k], H1[w+1, k]), H1[w, k]) \dots (22)$$

但し、 $(w-1)$ 、又は $(w+1)$ が存在しないときは、 $H2[w, k] = H1[w, k]$ 第 1 段階 (Step1) において、 $H1[w, k]$ は、単一の、あるいは孤立した 0 のバンドを無くした $H_n[w, k]$ であり、第 2 段階 (Step2) において、 $H2[w, k]$ は、単一の、あるいは孤立した突出したバンドを無くした $H1[w, k]$ である。このようにして、上記 $H_n[w, k]$ は、 $H2[w, k]$ に変換される。

【0076】次に、上記時間軸方向へのフィルタ処理に

$$H_{\text{speech}}[w, k] = 0.7 \cdot H2[w, k] + 0.3 \cdot H2[w, k-1] \dots (23)$$

また、背景雑音の信号に対しては、次の (24) 式に示すような時間軸での円滑化、あるいはスムージングを行

$$H_{\text{noise}}[w, k] = 0.7 \cdot \text{Min_H} + 0.3 \cdot \text{Max_H} \dots (24)$$

この (24) 式において、 Min_H 及び Max_H はそれぞれ、

$$\text{Min_H} = \min(H2[w, k], H2[w, k-1])$$

$$\text{Max_H} = \max(H2[w, k], H2[w, k-1])$$

にて求められる。

また、上記関係を示す曲線は、 $NR[w, k]$ (図 7 においては、 NR_w で示される) 値に応じて、範囲 L で変化するが、各値における曲線は、 $NR[w, k]=1$ の場合と同様な傾向で変化することがわかる。

【0073】また、フィルタ処理部 8 では、上記 $H_n[w, k]$ 値が周波数軸方向と時間軸方向とについて円滑化するフィルタ処理を行い、得られる信号として円滑化信号 $H_{t_smooth}[w, k]$ が出力される。上記周波数軸方向へのフィルタ処理は、信号 $H_n[w, k]$ の有インパルス応答長を短くする効果がある。これにより周波数領域での乗算によるフィルタの実現に起因する環状畳み込みによるエリアシングの発生を未然に防いでいる。また、上記時間軸方向へのフィルタ処理は、突発的な雑音を抑えるフィルタの変化の速さを制限する効果がある。

【0074】先ず、上記周波数軸方向へのフィルタ処理についての説明を行う。上記各バンドの $H_n[w, k]$ に、メディアン (中央値) フィルタ処理が施される。次の (21) 式及び (22) 式にて、この方法を示す。

【0075】

30 は、 $H1[w, k] = Hn[w, k]$

についての説明を行う。この時間軸方向へのフィルタ処理を施す際において、入力信号には、音声 (speech)、バックグラウンドノイズ、そして音声 (speech) の立ち上がり部分である過度的状態の 3 種あることを考慮に入れる。音声の信号 $H_{\text{speech}}[w, k]$ に対しては、次の (23) 式に示すように、時間軸での円滑化、あるいはスムージングを行う。

40 【0077】

う。

【0078】

【0079】また、過度的状態の信号に対しては、この時間軸でのスムージングを行われない。

【0080】以上のスムージング処理が行われた信号を用いて、(25) 式により円滑化出力信号 H

50 $t_smooth[w, k]$ を得る。

【0081】

$$H_{t_smooth}[w, k] \\ = (1 - \alpha_{tr}) (\alpha_{sp} \cdot H_{speech}[w, k] + (1 - \alpha_{sp}) \cdot H_{noise}[w, k]) + \alpha_{tr} \cdot H_2[w, k] \quad (25)$$

ここで、(25) 式中の α_{sp} は次の (26) 式から、 α_{tr} は次の (27) 式からそれぞれ求められる。

【0082】

【数16】

$$\alpha_{sp} = \begin{cases} 1.0 & SNR_{inst} > 4.0 \\ (SNR_{inst} - 1) \cdot \frac{1}{3} & 1.0 < SNR_{inst} < 4.0 \quad \cdots (26) \\ 0 & \text{上記以外} \end{cases}$$

$$\text{ここで } SNR_{inst} = \frac{RMS[k]}{MinRMS[k]}$$

$$\alpha_{tr} = \begin{cases} 1.0 & \delta_{rms} > 3.5 \\ (\delta_{rms} - 2) \cdot \frac{2}{3} & 2.0 < \delta_{rms} < 3.5 \quad \cdots (27) \\ 0 & \text{上記以外} \end{cases}$$

$$\text{ここで } \delta_{rms} = \frac{RMS_{local}[k]}{RMS_{local}[k-1]}$$

$$RMS_{local}[k] = \sqrt{\frac{1}{FI} \cdot \sum_{j=FI/2}^{FI-FI/2} (y - framej, k)^2}$$

【0083】続いて、バンド変換部9では、フィルタ処理部8からの、例えば18バンド分の円滑化信号 $H_{t_smooth}[w, k]$ が、例えば128バンド分の信号 $H_{128}[w, k]$ に、補間処理により拡張変換され、この変換された信号 $H_{128}[w, k]$ が出力される。この変換は、例えば2段階で行っており、18バンドから64

バンドへの拡張はゼロ次ホールドにより、64バンドから128バンドへの拡張はローパスフィルタ型の補間処理により、それぞれ行っている。

【0084】次に、スペクトラム修正部10では、高速フーリエ変換処理部3で得られたフレーム化信号 $y - framej, k$ の高速フーリエ変換処理にて得られるFFT係数の実部と虚部とに各々上記信号 $H_{128}[w, k]$ を乗じてスペクトラム修正、すなわち雑音成分を低減する処理が行われ、得られた信号が出力される。この結果、スペクトルの振幅は修正されるが位相は変形を受けない。

【0085】次に逆高速フーリエ変換処理部11では、スペクトラム修正部10にて得られた信号を用いて、逆高速フーリエ変換処理が行われ、得られたIFFT信号が出力される。

【0086】次に、オーバーラップ加算部12では、各フレーム毎のIFFT信号のフレーム境界部分についての重ね合わせが行われ、得られた出力音声信号が音声信号出力端子14より出力される。

【0087】また、ノイズ低減装置に適用した本発明の

音声信号の雑音低減方法の他の例を図8に示す。なお、図1に示したノイズ低減装置の構成と共通する構成部分については、これら構成部分を図1と同一の番号で示し、動作説明を省略する。

【0088】上記ノイズ低減装置は、入力音声信号を周波数軸の信号に変換する高速フーリエ変換処理部3と、上記入力音声信号から雑音部分を除去する際のフィルタ処理のフィルタ特性を制御するHn値計算部7と、上記Hn値計算部7にて得られるフィルタ特性に応じたフィルタ処理により、上記入力音声信号から雑音を低減するスペクトラム修正部10とを有して成るものである。

【0089】ここで、上記Hn計算部7を有して成る雑音抑圧フィルタ特性生成部35において、バンド分割部4は、高速フーリエ変換処理部3から出力される入力音声信号を高速フーリエ変換処理して得られる周波数スペクトルの振幅値を、例えば18バンドに分割して、バンド毎の振幅 $Y[w, k]$ を、RMS、推定雑音レベル、最大SNR計算部31と雑音スペクトル推定部26と初期フィルタ応答計算部33とに出力する。

【0090】また、RMS、推定雑音レベル、最大SNR計算部31は、フレーム化処理部1にて出力される $y - framej, k$ とバンド分割部4にて出力される $Y[w, k]$ とからフレーム毎のRMS値 $RMS[k]$ 、推定雑音レベル値 $MinRMS[k]$ 及び最大RMS値 $MaxRMS[k]$ を算出し、これら値を雑音スペクトル推定部26及び $adj1$ 、 $adj2$ 、 $adj3$ 計算部32に出力する。

【0091】また、 $adj1$ 、 $adj2$ 、 $adj3$ 計算部32は、 $RMS[k]$ 、 $MinRMS[k]$ 及び $MaxRMS[k]$ に基づいて、 $adj1[k]$ 、 $adj2[k]$ 及び $adj3[w, k]$ を算出し、これら値をN値計算部6に送る。

【0092】また、初期フィルタ応答計算部33は、雑音スペクトル推定部26から出力される雑音時間平均値 $N[w, k]$ と、バンド分割部4から出力される $Y[w, k]$ とをフィルタ抑圧曲線テーブル部34に送り、フィルタ抑圧曲線テーブル部34に収納される $Y[w, k]$ と $N[w, k]$ とに応じた $H[w, k]$ の値を算出し、この $H[w, k]$ をHn値計算部7に出力する。なお、フィルタ抑圧曲線テーブル部34は、 $H[w, k]$ に関する表が格納されている。

【0093】図1に示したノイズ低減装置や、図8に示したノイズ低減装置にて得られた出力音声信号は、例えば携帯用電話機の各種エンコーダ回路や、音声認識装置の信号処理回路等に送られる。あるいは、携帯用電話機のデコーダ出力信号に本雑音抑圧処理を施してもよい。

30

40

50

【0094】また、図9及び図10は、本発明の音声信号の雑音低減方法により雑音抑圧されて得られる音声信号（図中においては、黒）及び従来の雑音低減方法により雑音抑圧得られる音声信号（図中においては、白抜き）の歪を表すものである。なお、図9は、入力音声信号を20ms毎にサンプリングしたセグメントに対してSN比を求めて、このセグメント部分における歪を見積もってプロットしたグラフである。また、図10は、上記セグメントに対してSN比を求めて、このSN比に対して得られる全入力音声信号の歪を見積もってプロットしたグラフである。縦軸は歪を表して、高いほど歪が小さく、低いほど歪が大きくなるものである。また、横軸は上記セグメント部分のSN比を表して、右へ行くほどこのSN比は高くなる。

【0095】また、図9及び図10によれば、本発明の音声信号の雑音低減方法により雑音抑圧されて得られる音声信号は、従来の雑音低減方法により雑音抑圧得られる音声信号と比較して、特にSN比が20を越える高いSN比のところで、歪が少なくなっている。

【0096】

【発明の効果】以上説明してきたように、本発明に係る音声信号の雑音低減方法によれば、第一の値及び第二の値を用いて、入力音声信号から雑音を除去するフィルタ処理のフィルタ特性を制御することで、簡単な構成にて上記入力音声信号の最大SN比に応じたフィルタ処理にて上記入力音声信号から雑音を除去する、特に高いSN比での上記フィルタ処理による音声信号の歪を小さくすることが可能になり、また、上記フィルタ特性を得るための演算量の削減を図ることが可能になる。

【0097】また、本発明によれば、上記フィルタ特性を制御するための第一の値を、入力信号スペクトルのレベルと推定雑音スペクトルのレベルとで構成される表を用いて算出することで、上記フィルタ特性を得るための演算量の削減を図ることが可能になる。

【0098】また、本発明によれば、上記最大SN比及びフレーム毎の雑音レベルに応じて得られる第二の値を用いて、上記フィルタ特性を制御することで、このフィルタ特性を得るための演算量の削減を図ることが可能であり、また、上記フィルタ特性による最大雑音低減量を上記入力音声信号のSN比に応じて変化させることが可能になる。

【図面の簡単な説明】

【図1】ノイズ低減装置に適用した本発明の音声信号の

雑音低減方法の第一の例を示す図である。

【図2】本発明の実施例におけるエネルギー $E[k]$ 及び減衰エネルギー $E_{decay}[k]$ の具体例を示す図である。

【図3】本発明の実施例におけるRMS値 $RMS[k]$ 、推定雑音レベル値 $MinRMS[k]$ 及び最大RMS値 $MaxRMS[k]$ の具体例を示す図である。

【図4】本発明の実施例におけるdB表示の相対エネルギー $dB_{rel}[k]$ 、最大SN比 $MaxSNR[k]$ 、及び雑音判別の閾値の1つである $dB_{thres_{rel}}$

10 $[k]$ の具体例を示す図である。

【図5】本発明の実施例における最大SN比 $MaxSNR[k]$ に対して定義される関数としての $NR_{level}[k]$ を示すグラフである。

【図6】本発明の実施例における $NR[w, k]$ と最大雑音低減量との関係をdB表示で示すグラフである。

【図7】本発明の実施例における $Y[w, k]/N[w, k]$ 比と $H_n[w, k]$ との $NR[w, k]$ に応じて得られる関係をdB表示で示すグラフである。

20 【図8】ノイズ低減装置に適用した本発明の音声信号の雑音低減方法の第二の例を示す図である。

【図9】上述の各ノイズ低減装置により雑音抑圧されて得られる音声信号のセグメント部分の各セグメント部分のSN比に対する歪を表すグラフである。

【図10】上述の各ノイズ低減装置により雑音抑圧されて得られる全音声信号の上記各セグメント部分のSN比に対する歪を表すグラフである。

【符号の説明】

3 高速フーリエ変換処理部

4 バンド分割部

30 5 雑音推定部

6 NR値計算部

7 H_n 値計算部

21 RMS計算部

22 相対エネルギー計算部

23 最大RMS計算部

24 推定雑音レベル計算部

25 最大SNR計算部

26 雑音スペクトル推定部

31 RMS、推定雑音レベル、最大SNR計算部

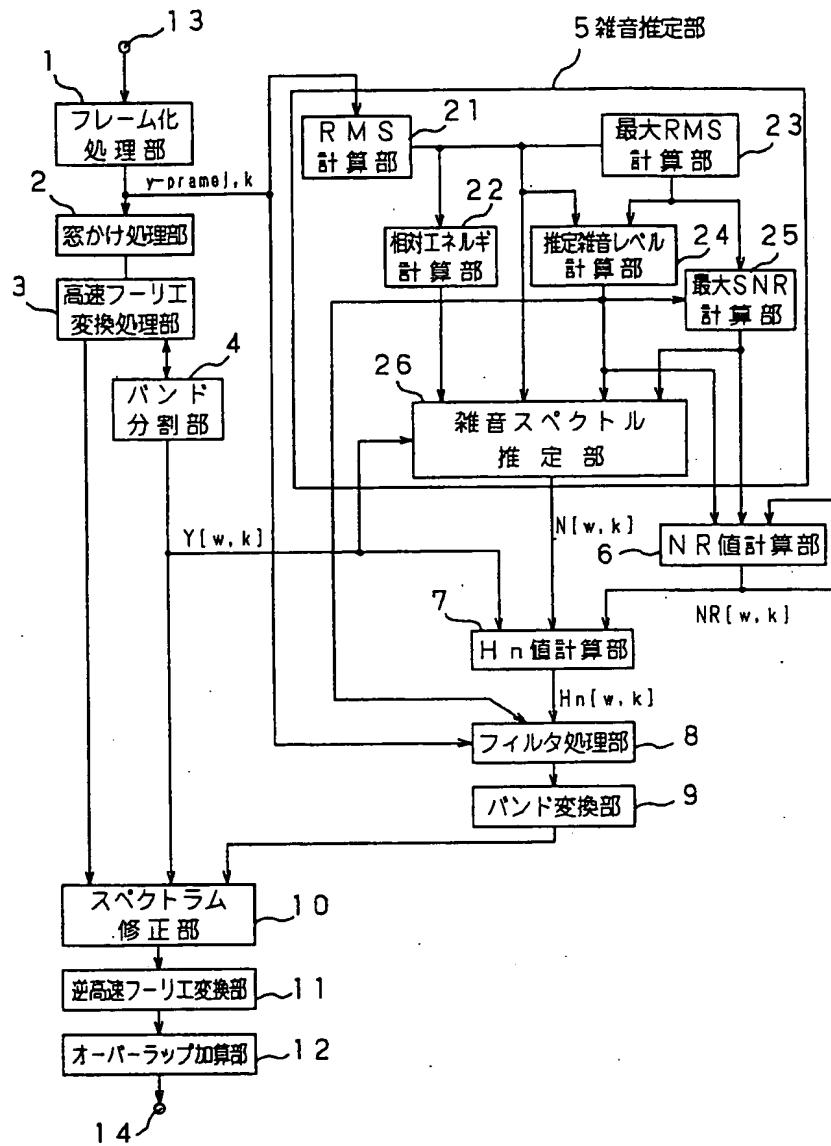
40 32 $adj1$ 、 $adj2$ 、 $adj3$ 計算部

33 初期フィルタ応答計算部

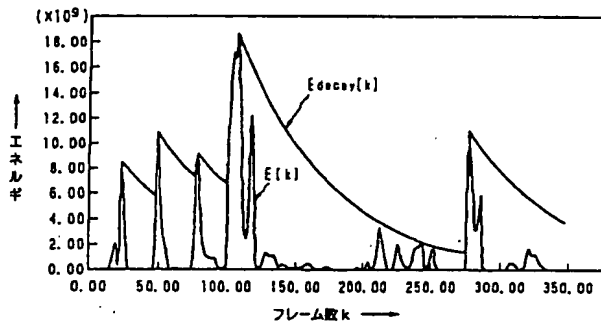
34 フィルタ抑圧曲線テーブル部

35 雑音抑圧フィルタ特性生成部

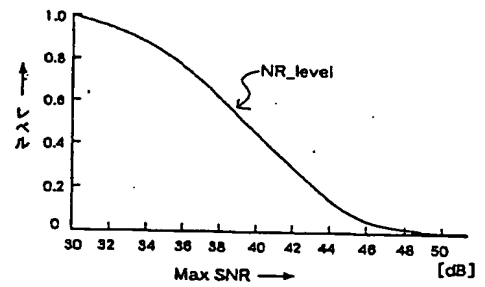
【図1】



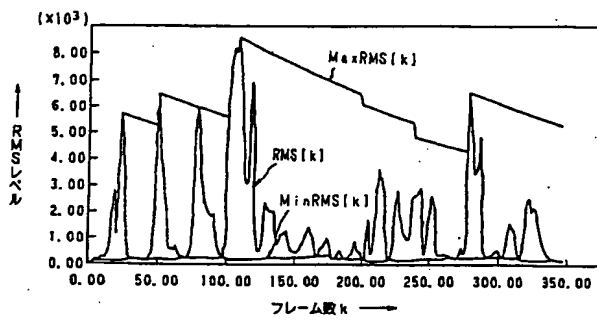
【図2】



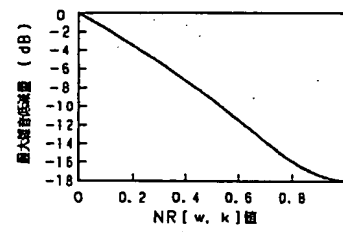
【図5】



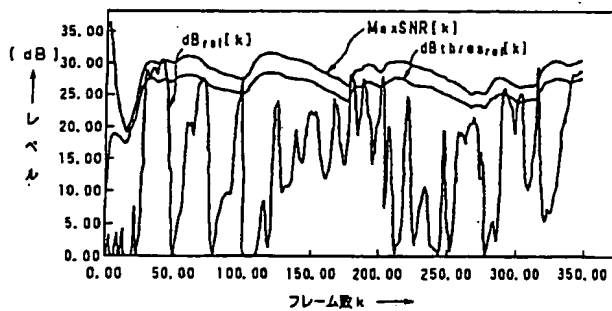
【図3】



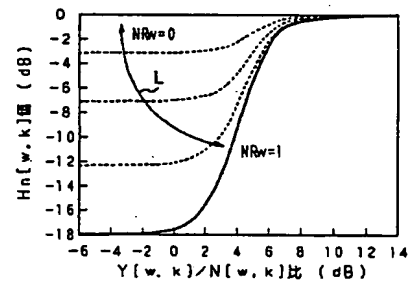
【図6】



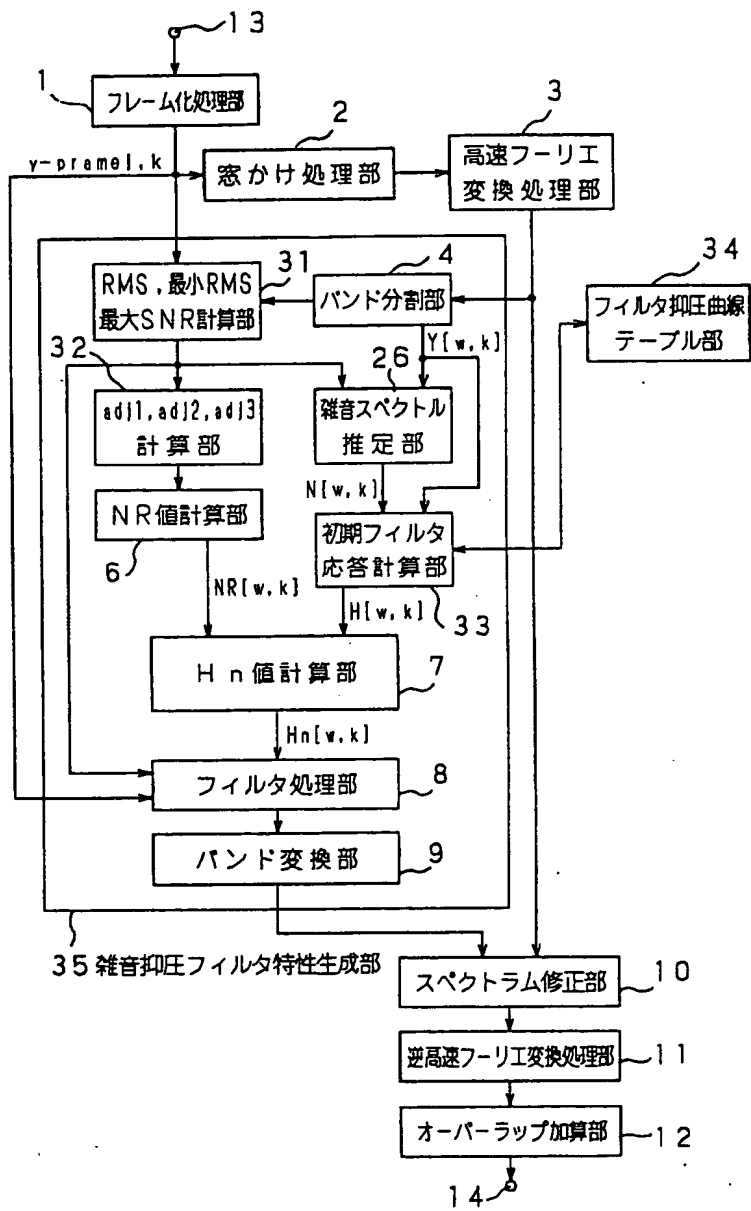
【図4】



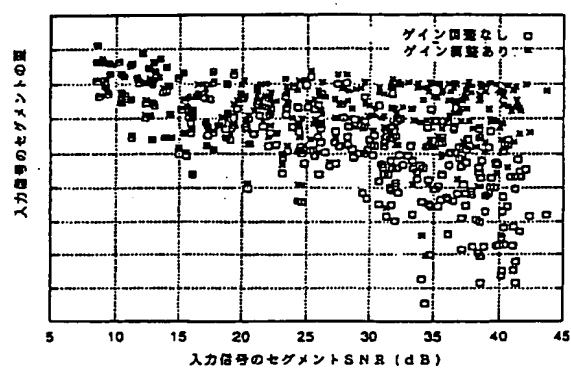
【図7】



【図 8】



【図9】



【図10】

